

8

The volume editors provide a comparative analysis of the evaluation approaches employed by each of the theorists relative to a number of important evaluation issues.

Unraveling Theorists' Evaluation Reality

Marvin C. Alkin, Christina A. Christie

The theorists in this volume have confronted the exercise presented to them in vastly different ways. As editors, we asked them “to consider how you would evaluate this program” and “to make and specify assumptions about the program context.” The primary task, however, was to describe “a course of action, which may have consequences that lead to other evaluation design decisions.” All of the theorists did indeed do this. There was great diversity in the evaluation approaches presented and, of course, in the presentation of approaches. As mentioned in Chapter Two, we anticipated diversity on both of these dimensions and were pleased to see that it emerged. Otherwise our experimental exercise would have been, dare we say, uninteresting. In this chapter, we explore the differences in approaches that are evident to us, as well as the similarities (which ought to be just as remarkable in some instances) on several dimensions of evaluation practice. We will comment first, however, on each theorist’s overall approach to the task before moving on to the discussion of comparing approaches, and have included a visual summary at the end of the chapter in Table 8.1.

Tackling the Exercise: Overall Format and Approach

Each of the four theorists presented his or her evaluation proposal creatively—describing to the reader what it would be like to walk through an evaluation “in his or her shoes.” What would we encounter, and how would we deal with it? Henry’s approach was relatively straightforward. He made specific assumptions about what stakeholders wanted out of the evaluation and then proceeded to establish a design. Greene and King discussed at

length the kinds of interactions that would take place and processes to be implemented, which would lead to further evaluation activities. And Donaldson engaged the reader as an interviewee for a position on his evaluation team. This mixture of format is what we had hoped for. We believe that the way theorists chose to portray their approach to evaluation, in fact, provides insights into their way of thinking about and approaching evaluation.

Theorists' Assumptions and General Evaluation Process

The case scenario in Chapter One was designed to provide answers to many of the potential questions an evaluator may have about the program but obviously left many unanswered questions. The absence of interaction with real-life stakeholders required that each theorist make assumptions, both implicit and explicit, about the case: its stakeholders, context, resources, and the like. In general, theorists made assumptions that moved the evaluation along in a direction that would accommodate the theorists' proclivities. Availability of resources for the evaluation serves as a good example of the kinds of key assumptions made.

Theorists made pivotal assumptions about both the financial and human resources available for the conduct of their evaluation. Henry, for example, assumes that resources will be secured to conduct a multisite evaluation of the program model. Donaldson, in contrast, explicitly states that he assumed that resources were limited and designed his evaluation accordingly, even commenting that if resources were assumed to be unlimited, he would have designed the evaluation differently. Greene also created a design mindful of monetary restrictions. King does not assume that limitless financial resources are available; however, she does make the assumption that stakeholders will be highly committed to and involved in the evaluation. This assumption necessitates a significant human resource commitment.

Obviously, the theorists' assumptions went beyond what financial resources are available. Many assumptions were also made about which stakeholders will be contacted, stakeholders' responses to theorists' proposals about actions to be taken, and, in some instances, the consequences of those actions. Our analysis of each theorist's approach will bring to bear some of the other more important assumptions made.

Theorists' Approaches

In this section, we first summarize the evaluation designs presented by each theorist, pointing to some of the unique aspects of that theorist's presentation. Following this discussion, we present the particular themes that emerged to us across theorists' designs. In some instances, these themes highlight commonalities in presentations. Conversely, a theme may be developed to distinguish theorists' presentations.

Jennifer C. Greene. Greene proposed a “value-engaged approach.” She noted that this approach is derived from Stake’s concern for responsiveness, but is supplemented with values engagement drawn from democratic and culturally responsive traditions in evaluation. In her chapter, we note the familiar features of evaluation that are present. First, she attempts to develop an understanding of the context and the program to be evaluated. She then works with a variety of stakeholders to identify key evaluation questions and sets up a procedure for establishing criteria for making judgments of program quality. Next, Greene established an evaluation design that incorporated multiple and mixed methods for acquiring information and attended to the process of conducting the evaluation in ways that influenced an evolving evaluation design. Greene did not comment specifically on evaluation reporting, but presumably it would involve a combination of written reports and “meaningful dialogues with stakeholders.” Finally, she proposed conducting a meta-evaluation. One of our colleagues, Lynn Winters, assistant superintendent for planning and evaluation in the Long Beach (California) Unified School District, on reading the chapter, suggested that the approach reminds her of many evaluations that she has seen: “I would like to know what this [value-engaged] approach really means in terms of perspective and how it affects methodology and the kinds of questions one would ask.”

We acknowledge Winters’s point. Nevertheless, the attention to broad-based stakeholders, while it might be common in many approaches, is substantially more intensive in what Greene proposes. Indeed, the evaluation priorities emerged from discussions with parent communities as well as from a discussion at a faculty meeting. Moreover, the criteria for making judgments of program quality were established through discussions with diverse stakeholders. As noted in one of Greene’s memos, initial specification of criteria for making judgments emerged not only from the school community, but also from parents and families. It is perhaps difficult to envisage in a written evaluation design the sensitivity of the evaluator to principles of social justice that undergird each activity.

And so we raise the issue about how possible it is to judge differences in theoretical applications by the written document that is produced. To a large extent, it is necessary to observe Greene as she engages in the evaluation, who she listens to, what special efforts she puts forth, and so forth.

Gary T. Henry. Henry’s approach is concerned with identifying the mechanisms that lead to program outcomes. He focuses on the efficacy of the overall Da Vinci program model, and the evaluation he proposes is designed to do just that. There are no surprises in his design. He offers a well-thought-out quasi-experimental study. In order to conduct this study, Henry assumed that resources could be secured to support this large-scale design, which includes twenty-five Da Vinci model schools and twenty-five comparison schools. Although he does not detail the budget for his study, he does suggest offering each comparison school \$10,000 for participation

(totaling \$250,000). And so it is important to note that Henry's design would likely far exceed the cost of implementing any of the other theorists' designs.

Henry convenes an evaluation advisory group consisting of representatives from the evaluation funder, the state educational office, the state superintendent's office, and other high-level organizations to advise the evaluation and aid in the interpretations of findings. It is apparent from the selection of advisory board members that Henry's evaluation is not driven at the local school level. He does, however, suggest including local school stakeholders to identify the values of those affected by the program, and uses sampling techniques to ensure representation at levels. At the initial stage of the evaluation, which he refers to as values inquiry, he offers both qualitative and quantitative methods for identifying stakeholders' most highly valued program outcomes. He then develops a program theory, with the assistance of the Da Vinci program developers and teachers systematically selected from four schools, that focuses on the outcomes identified during the values inquiry stage. Outcomes are measured by extant achievement measures, and program implementation is measured using multiple data sources. Henry is the only theorist to describe in detail the sampling methods and data analysis techniques that he would use. This is seemingly because Henry's approach relies heavily on the success of the implementation of the study design and data collection procedures.

Henry was the only theorist to state that one of the explicit purposes of the evaluation was to assess the merit and worth of the program. Winters points out that the strength of Henry's chapter is that he was "proactive about focusing the evaluation on the Da Vinci Model. . . . He saw through . . . to the 'core beast' needing review . . . [for] useful policy decisions." We recognize the concern about the evaluation of the Bunche–Da Vinci model but would like to note that by moving in that direction, Henry bypasses the immediate concern for satisfying local concerns. His concern for generating information for policy decisions for the State of Columbia is, however, what differentiates the focus of his evaluation and, subsequently, his evaluation approach.

Stewart I. Donaldson. Donaldson's application of program theory-driven evaluation science as it is described in his chapter has three steps: (1) develop a program theory, (2) identify evaluation questions, and (3) design and conduct the evaluation. Donaldson tells us that the first two steps should take about three months and that he would contract for steps 1 and 2 before contracting to conduct step 3.

Overall, Donaldson describes his role as one of a facilitator of discussion. He first facilitates a discussion among stakeholders to establish a program theory. Donaldson described this as an interactive process of making stakeholders' assumptions and understandings of the program explicit—that is, they develop a program theory. He does not describe for us the specific group of stakeholders that would participate in this process. Nevertheless, he assumes that at Bunche–Da Vinci, the program theory is

anchored in student performance—even though he anticipates that some stakeholders may object to this. The plausibility of the program theory is then assessed using extant literature and revised accordingly. Evaluation questions are formulated that would focus on curriculum implementation, program operations and educational service delivery, and program outcomes. Although stakeholders' opinions about the value of each question are considered, it seems that Donaldson, while cognizant of the program theory, makes the final decision about which questions to pursue.

In Donaldson's evaluation, methodological choices are informed by the program theory. He speaks of the evaluator as a facilitator of a process in which the evaluator educates stakeholders about the benefits and challenges of using particular methods and leads to agreement about which data sources and collection techniques will be used. He also stresses the importance of establishing criteria of merit with stakeholders to justify conclusions and recommendations and to increase use.

Shadish, Cook, and Leviton (1991) describe theory-driven evaluation (some say too generously) as a comprehensive "third-generation" approach to evaluation, one that incorporates methods from previous evaluation theories. When reading Donaldson's chapter, one cannot help but detect methods from other theoretical approaches. For example, his attention to García and Chase, the program administrators, reminds us of the work of user-oriented evaluation theorists who express concern for developing evaluations attentive to the needs of decision makers. Donaldson offers the Program Evaluation Standards as a means for establishing and checking evaluation processes. There is also direct reference to establishing criteria of merit, a process that Scriven argues is an essential component of any evaluation. Of course, there is the focus on establishing the connection between program outcomes and processes, which Campbell and Cook argue is the purpose of evaluation. Finally, Donaldson describes the evaluator as a facilitator of stakeholder beliefs and values, reminiscent of democratic approaches. Thus, Donaldson's approach could be seen as an eclectic mix of procedures and processes, formulated around a program theory.

Jean A. King. King's chapter is particularly significant in the way in which it demonstrates the thinking process of an evaluator. In King's case, we have an evaluator who, as a consequence of being concerned with evaluation utilization, is focused on building evaluation capacity as an end result of the evaluation process. And how is capacity built? At its core, capacity building involves the creation of structures for attaining the active involvement of school-based participants. This means that active learning (about evaluation) must take place: committees are engaged in deciding what is to be done (an advisory group). King structures activities to build an evaluation infrastructure. This is done through systematically seeking to understand the context. Various direct activities to enhance infrastructure are proposed, including having staff participate in a small-scale, but nevertheless visible, participatory inquiry project.

But in discussing the activities proposed by King, we clearly have put the cart before the horse. King proposes a great deal of upfront work before she even agrees to do the evaluation. She engages in a self-evaluation to determine whether the evaluation situation is a good fit for her own skills, interests, and background, and then some “reconnoitering” and informal research are required. She wants to know about the specific context and to some extent about comparable programs. All of this precedes the negotiation of the contract, in which she sets specific conditions to be fulfilled before agreeing to participate.

One of the conditions that King mandates is a substantial (perhaps very substantial) amount of active involvement by school personnel. We wonder about the prevalence of situations where such involvement is possible. Does insistence on active involvement mean that a school like Bunche–Da Vinci would not obtain King’s services, and she would do evaluations only in settings where teachers are not so overwhelmed?

Building evaluation capacity takes time. This is quite evident in King’s description of the evaluation process. It is indeed worthwhile and rewarding to not only provide information on the success of a program, but also to create a structure in which people respect and relish evaluation information and continually engage in the process as a means of improving their program. Building an evaluation community is hard work. Do the parties involved in Bunche–Da Vinci have the patience to wait for the answers they seek? Perhaps many evaluation situations are not instances for potential evaluation capacity building.

King addresses some of these issues in Chapter Seven. She expresses optimism about the potential for completing the evaluation. We wondered, however, whether there are lost opportunities, that is, evaluations in which she chose not to participate. Does her stance reflect the autonomy of a university-based evaluator and not that of an evaluator in full-time practice? She, like us, has the luxury of deciding what evaluation studies she would do in addition to her university employment. Are there some markers that she uses at the outset of the project that will allow her to make early decisions as to whether this is an evaluation in which she could or should be involved?

Common Themes

Of interest to us was the extent to which theorists’ approaches were similar. For example, some of the steps that Greene and King proposed with respect to engaging stakeholders are particularly comparable. However, the intent behind the action, as described by the theorist, is different. Acknowledging this, we engaged in a lengthy discussion about the extent to which a theorist’s intention behind an action is, or is not, evident in the action itself. Let us explain further.

We understand that the intent that motivated Greene and King to engage stakeholders was different because they told us so. We asked them to explain what they would do and why—and they did. King wanted to build evaluation capacity by engaging stakeholders in the process of conducting the evaluation. Greene wanted to engage a broad spectrum of stakeholders in the evaluation process in order to be responsive to their values. In our view, there was a difference in intensity in the engagement of stakeholders. Perhaps King's approach involved more "doing" and Greene's more thinking, judging, and communicating. But in any case, imagine being a stakeholder who is unfamiliar and unconcerned with evaluation terminology or theoretical nuance. The action, absent a description of what it is intended to yield, looks much the same. So we ask the question, When similar actions or practices are motivated by different intents, that is, if King and Greene engage stakeholders in a similar fashion but do so to accomplish a different end, how can one be guaranteed that King's action produced what was intended—and indeed not what Greene intended it to produce instead? This observation may help to explain why in Christie's study (2003) of evaluation theorists' practice, some theorists with seemingly different theoretical approaches, when asked to describe only their practices (absent an explanation of intent or motivation), quantitatively look similar.

On Theory. Each theorist was very consistent with his or her own purported theoretical position—perhaps too much so. Again, the process of being selected to write a chapter may have led the theorists to believe that they were commissioned to portray how an evaluation would be conducted from the vantage point of their specific theoretical position. Indeed, each of the chapter authors commented on his or her perspective on evaluation as a part of the evaluation proposal. Perhaps this is not surprising. The orientation of an evaluator sets the context for what an evaluator proposes and how he or she will do it. Thus, we find little to comment on with respect to adherence to theoretical position. To the extent to which the context matched his or her own theoretical position, each theorist proposed an evaluation along the lines of that position.

While the theorists chosen for participation in this exercise represented each of the branches of our evaluation theory tree (Greene, values; King, use; and Donaldson and Henry, methods), they are not, and could not be, representative, in any sense, of all theorists who might be depicted on the branch of the tree on which they have been placed. We view Greene as a "democratic evaluator." Yet there are differences between her approach and, say, those of Stake (2003) or House (House and Howe, 1999). We see the extent of her desire for substantive inclusion of underrepresented individuals in all stages of the evaluation as greater than that of Stake. Furthermore, we believe that the extent to which her inclusion principles would extend to advocating for groups in the absence of broad participation would be less than that which House (2003) might employ. Obviously there are vast differences in

the way that she conceives of evaluation from what Scriven (also concerned about valuing) would propose (2004). To give one further example, King and others (for example, Cousins and Whitmore, 1998; Preskill and Torres, 1999) on the upper reaches of the use branch who are strongly engaged in evaluation capacity building differ from other use-oriented evaluators such as Stufflebeam (2003) and even Patton (1997). (We cannot help but note also that on the face of it, there are similarities that go across branches. One theorist can generally agree with another's procedures and actions, but the intentions behind the actions are the basis for theoretical nuances that make a difference.) Thus, we are not attempting to demonstrate differences in evaluation practice between broad theoretical perspectives. Rather, this volume describes four theorists' unique perspectives and how, from that perspective, they would engage in the conduct of an evaluation, given the particular context of the case scenario presented.

Stakeholder Engagement. Each of the theorists in this volume addresses the need for stakeholder involvement. Indeed, by now, this has become an implicit part of almost all evaluation theories (Christie, 2003). Yet there are substantial differences between theorists in the choice of stakeholders to be included, the stages at which they participate, and the nature of their involvement. King's stakeholders are primarily Bunche–Da Vinci faculty and staff. While others' inputs are reflected in the process, the focus in evaluation capacity building for King is the primary users, primarily staff and teachers. These users are engaged as primary “doers” at every stage of the evaluation. They are not simply consulted to provide input, and they are not stakeholders in the sense of having their views reflected and advocated by the evaluator. Instead, their stakeholder role is active engagement in all phases of the evaluation.

Greene also engages stakeholders substantially at most phases of the evaluation. Her stakeholder group is broad based, with particular attention to the inclusion of people traditionally underrepresented or, in the absence of sufficient participation, to the reflection of their views.

Donaldson engages only García and Chase in his discussion of the evaluation, although he talks specifically about having them identify leaders of key stakeholder groups for him to contact to describe the evaluation plan. There is no discussion, however, of the role of these additional stakeholders in the evaluation process. Beyond Chase and García, we never learn who these stakeholders are in the case of Bunche–Da Vinci or exactly how their roles are envisioned.

Henry is concerned with having broad stakeholder input, particularly at the values inquiry stage of his evaluation. Broad-based stakeholder participation, however, is not maintained throughout the evaluation process. In fact, it seems to taper as Henry moves through his evaluation plan. That is, the most significant level of stakeholder participation occurs at the beginning stages, including program theory development, but interpretation of findings is left to the evaluator and the evaluation advisory committee. Of

course, this advisory committee comprises individuals with a vested interest in the program, but they represent just a small segment of possible stakeholder groups.

Use. Use has become a central theme in the evaluation theory literature. There is general agreement that information yielded from evaluations is intended to be used. There is debate, however, about the extent to which potential evaluation use should serve as a driving force behind an evaluation's design and implementation. Nonetheless, most theorists suggest that evaluators include stakeholders in the evaluation process in order to increase the odds that the evaluation will be used.

Thinking about stakeholder involvement in the evaluation process as a means for increasing use, we distinguish between depth and breadth of involvement (House, 2003). When referring to depth of involvement, we are considering the extent to which stakeholders participate in the process; when referring to breadth, we are considering the number of stakeholders and constituencies reached during the evaluation process. In our experience, often one is sacrificed at the expense of the other. That is, it is very difficult to have both great depth and breadth of stakeholder participation in the evaluation process, and, as a result, the evaluator often has to choose between the two. This choice, however, is often determined from the outset by the overall evaluation approach.

The theorists in this volume, by and large, did not address specifically what they would do with respect to use. It is our belief that the task presented to the theorists (to describe how they would study a particular program) did not lend itself to an in-depth discussion about use. And so we surmise that this is a function of the task rather than a conscious decision on the part of theorists to not attend to the topic in depth. Thus, we will address use as it related to stakeholder involvement in the evaluation process.

King addressed use most explicitly, which is what one would presume based on the nature of her approach. She is entirely focused on use, but not primarily the use of specific evaluation findings. Rather, her concern is with what is referred to in the literature as process use (Patton, 1997). Focusing on the small group of stakeholders within the school, she seeks to engage them in the process of evaluation to build organizational capacity. Her approach involves deep stakeholder involvement.

Greene's approach involves somewhat less depth of stakeholder involvement but considerably greater breadth. The extensive inclusion of a wide range of stakeholders, while not focusing specifically on use, presumes that a higher quality of use will occur—one that reflects values of social justice.

Relative to the other theorists, Henry proposes to involve the largest number of stakeholders. As such, he is concerned with breadth of stakeholder involvement and broad use of his findings. Use of information at the local level is important, but is offset by a value for generating information to be used to inform policy decisions. Donaldson involves stakeholders at the local level, but also mentions involving stakeholders beyond those identified

in the case. Although not stated explicitly, we presume that Donaldson is concerned with use beyond the program level, for example, is interested in generating information for the Da Vinci Learning Corporation and the school district, but is less concerned in this particular evaluation context with providing for state policy decisions (like Henry).

Program Theory. All the theorists in this volume acknowledge the importance of understanding program theory as part of the evaluation. However, the emphasis placed on the role of program theory in the evaluation process is vastly different. This is not surprising given each theorist's primary focus.

Greene is mainly concerned about value-engaged evaluation; her focus is primarily on the inclusion of the values and interests of underrepresented groups. Henry expresses concern for performing evaluation in policy contexts—which usually requires the use of experimental or quasi-experimental designs to be considered appropriately valid. Thus, program theory is important but secondary to design concerns. King, in her attention to building evaluation capacity, is more concerned with the individuals and the process in which they engage. Clearly Donaldson's primary focus is program theory as the route for understanding the relationship between and impact of program activities.

The cornerstone of Donaldson's approach is the development of a program theory. Once developed, the program theory prescribes and orders the questions that the evaluation will address. The evaluation is then built around the questions yielded from an analysis of the program theory. Winters notes that "Donaldson convinces me 'assumption-examining' is an important step in the evaluation design and is a perspective that doesn't show up explicitly anywhere else but drives all of the activities in other chapters." Indeed, Greene and Henry in particular acknowledge the necessity of understanding a program's theory. Donaldson's approach, however, moves beyond just understanding a program theory by defining program goals and testing the relationship of proposed activities to those goals.

Henry uses the program theory as a means for determining the connections between program processes and outcomes. He is most concerned, however, with examining program outcomes, and so it is not necessary to connect each program outcome to a program process. Different from Donaldson, Henry identifies the primary questions for the evaluation prior to program theory development. The program theory is used to help the evaluator determine where to look when explaining program outcomes.

Program theory is not the focal point of Greene's approach to evaluation. Yet a concern for "understanding the conceptual rationales" of the Bunche–Da Vinci is a part of her approach. Her strong attempt at understanding the school context adds light to her consideration of the program's theory.

While program theory is not central to the work of King, she nonetheless acknowledges its importance in some of the activities in which evaluation

participants engage. For example, in her chapter discussion related to making sense of test scores, she notes that the committee engaged in this activity "might develop program theory that would plan backward from the necessary achievement outcomes to identify explicit strategies to increase learning in specific areas."

Social Justice. We would argue that most professionals conducting evaluations of social and educational programs think of themselves as service providers and hope to be promoting justice, equity, and social betterment through their work (Donaldson and Christie, forthcoming). Some evaluators, however, pursue this (lofty) goal more intentionally through their actions. These approaches are sometimes described as social justice approaches. Christie (2003) found that those concerned with intentionally promoting social justice through their evaluation actions can be differentiated. From one perspective, the evaluator intends to promote social justice by increasing representation (House, 1993). From another perspective, social justice can be promoted by empowering those involved with the evaluation (Fetterman, 1996). Social justice approaches to evaluation are distinguished, however, from other approaches that pursue social betterment. All of the theorists in the volume expressed or implied a general concern for promoting social betterment. All were highly motivated by the desire to provide information that would improve the education of students. There are differences, however, in the ways in which this concern is addressed throughout the evaluation process and the extent of emphasis on social justice.

The title of Henry's chapter, "In Pursuit of Social Betterment," reflects his concern for conducting evaluations that contribute to the social good. He does not, however, intend to conduct a social justice evaluation as described by House or Fetterman, although his proposal to involve a broad group of stakeholders and the methods proposed for doing so (sampling, surveying, and focus groups) reflect, at least in part, some of the ideas put forth by House (for example, in House, 2003). Donaldson indicates a desire to implement a place that could "help them improve the way they educate students" (again, social betterment, but not particularly social justice). King also is clearly focused on social betterment rather than social justice as an agenda. She, like the others above, is implicitly concerned about issues related to social justice. However, her focus is social betterment—specifically through building the evaluation capacity of the school. Greene undoubtedly has the strongest social justice agenda in her design. She follows a value-engaged approach, which aspires to meet the educational needs of people traditionally underrepresented: racial and ethnic minorities, low-income people, and others.

Methods. All of the theorists proposed a mixed-method approach to studying Bunche–Da Vinci. Some were more specific about which methods they would use and when. Henry, for example, proposed using focus groups to identify the stakeholder values during the values identification phase of his evaluation. He said that he would later use observation methods to study

program implementation. Nevertheless, the data collection methods used to study program outcomes are primarily quantitative. As mentioned previously, Henry is the only theorist who described a specific data analysis plan. This description provides a rationale for the data collected as well as an understanding of how program success will be determined.

We found it interesting that both Henry and Donaldson stated explicitly that their approaches are “method neutral.” And this may be because their particular approaches are at times characterized as being primarily quantitative. This in fact was the case in the evaluation designs proposed in this volume. Nevertheless, it may be important for both Henry and Donaldson to state that their methodological choices are not determined *a priori* and that they are not limited to quantitative methods. But as evidenced by the proposals offered to evaluate Bunche–Da Vinci, each of their approaches stresses the importance of measuring program impact in a way that lends itself to the use of quantitative measures. In fact, it is our opinion that neither Henry nor Donaldson would consider examining Bunche–Da Vinci’s program outcomes or impacts qualitatively, that is, to simply describe a program outcome, without some statistical evidence of its impact. Both, however, propose using qualitative methods to measure program fidelity and implementation.

Greene’s design has some quantitative methodological components, primarily related to outcome measures, but it rests most heavily on qualitative methods. There is a strong focus in her evaluation work on discussions, conversations, classroom and school observations, and interviews with both staff and parents.

King’s methodology can perhaps be best described as participatory engagement designed to obtain skills and build capacity. There is some quantitative analysis (engaged in by program participants conducting a mini study). However, the methodology of qualitative approaches dominates the design.

Table 8.1 presents a comparison of the theory chapters in terms of the dimensions discussed in this chapter.

Last Word

We view the theorists’ chapters in this volume as a set of case examples. Thus, our interpretation of these cases is subject to validation of some type. As Yin (1989) notes, “A major way of improving the quality of case studies . . . is to have the draft cases reviewed by those who have been the subject of the study” (p. 144). We therefore invited the authors to provide their “last words” on the topic; most accepted the invitation.

JENNIFER C. GREENE: Evaluation is a complex social practice with multiple, interwoven strands. One strand represents the technical aspects of our

Table 8.1. A Comparison of Approaches

Stakeholders: Depth of Participation							
Approach	Stakeholders	Participation	Use Concern	Program Theory	Social Justice	Methods	
Greene	Value engaged	Broad based, attention to inclusion of traditionally underrepresented people	Very substantial engagement at most phases	Use reflects values of social justice	As an aid in understanding the conceptual rationales of the program	Social justice	Mixed methods—heavily qualitative
Henry	Quasi-experimental	Broad stakeholder input	Moderate participation at beginning stages	Information used to inform policy decisions	As an aid in understanding cause-and-effect relationships	Social betterment	Primarily quantitative
Donaldson	Program theory evaluation science	Program management	Engagement primarily at early stages	Primarily used at the program level	Route for understanding relationship between and impact of program activities	Social betterment	Primarily quantitative
King	Building evaluation capacity	Primary users: staff and teachers	Active engagement in all phases	Process use	As an aid to stakeholder planning	Social betterment	Mixed methods—primarily qualitative

work. To be good evaluation technicians, we use our expertise in social inquiry methodology. The ways in which we craft an evaluation design, select methods and samples, analyze and interpret data, and fashion evaluation reports all represent the quality of our technical knowledge. The evaluations presented in this volume well represent this technical dimension of our work, even as they offer varied perspectives on it. Another strand of evaluation practice is substantive and contextual. We endeavor to understand the quality of a given human endeavor in the specific form it takes in the particular contexts being evaluated. This strand of our work requires some conceptual and experiential familiarity with the evaluand and the contexts at hand. In this volume, the substance of relevance is schoolwide educational reform as packaged by the Da Vinci Learning Corporation *and* as implemented in a particular underresourced urban context. The editors of the volume appropriately noted that the four evaluators have differential conceptual and experiential familiarity with schoolwide educational reform in underresourced public school settings, as this indeed matters to the quality and influence of our work.

A third strand of evaluation practice is sociopolitical. It concerns the positioning of evaluation with respect to the program and policies being evaluated and relates directly to intended evaluation purposes and audiences. This strand is among the most contested in evaluation, as it invokes the multiple interests of diverse stakeholders, all of which are legitimate but some of which can be competing, even conflicting. The different evaluations presented in this volume clearly differ in terms of whose interests are addressed and which stakeholder concerns are privileged.

A final strand of evaluation practice is relational. It pertains to the evaluator's presence in the context being evaluated, to the roles and identities she takes on, to her interactions and communications with others in that context. In many ways, the other strands of evaluation are enacted through such roles, communications, and relationships; thus, the relational strand is a powerful one. I endeavored to share the relational strands of my envisioned evaluation in the Bunche–Da Vinci context through memos that referred to evaluative gatherings and events *and* that conveyed particular norms, notably of respect and reciprocity. Yet I fully appreciate the editors' perceptive insight that a linear written text cannot really capture this relational strand of evaluation, as it takes meaningful form only in action.

STEWART I. DONALDSON: Why have Alkin and Christie passed up the golden opportunity to do what this volume is all about: evaluate? As they pointed out, Shadish, Cook, and Leviton (1991) evaluated the various theories of evaluation practice they critically examined, and theory-driven evaluation emerged in a very favorable light. I was secretly hoping to find my plan for evaluating Bunche–Da Vinci to be ranked number 1 out of 4, and dreading the possibility, and ready to counter (or employ ego defense mechanisms), if it was ranked anything less. If they do not want to evaluate us themselves,

how about submitting these plans to a panel of experts, principals, teachers, or parents for judgment?

I suggest summative evaluation jokingly in my last word, where it is safe (I think) to assume they will not get any ideas along these lines, to help underscore the point that this exercise was not about which plan is the best. In fact, I think the answer to the questions above is that this exercise and volume is not about evaluating merit, but rather a good example of creative, descriptive research on evaluation theory and practice. Alkin and Christie's collegial approach to this exercise and research on evaluation theory and practice more generally (for example, Alkin and Christie, 2004; Christie, 2003) is refreshing, exemplary, and sorely needed. In my evaluation, these works make significant contributions toward advancing our understanding of the emerging discipline and profession of evaluation.

But let me end by evaluating a few of their observations about my plan for helping Bunche–Da Vinci. The discrepancies noted below may be due more to the lack of detail in my chapter, or my ability to communicate well in this format, than to errors of analysis. First, I give them an A for describing my approach as “an eclectic mix of procedures and processes, formulated around a program theory.” Engaging diverse stakeholders in an effort to fully understand their program (the evaluand), and then using that shared understanding to present evaluation options based on the best procedures, process, and methods available today (from across approaches and theories of practice) sums up my approach well. In contrast, Alkin and Christie's observations about stakeholder involvement and that the evaluation team would determine the evaluation questions made me realize I did not state clearly enough that I would engage as many stakeholders as made sense in the light of human resource and financial constraints and that the evaluation clients (stakeholder representatives), not the evaluation team, would ultimately decide which questions to pursue.

I was intrigued that both Gary T. Henry and I used the “method-neutral” defense, and by the bold statement that we would not consider qualitative methods for assessing Bunche–Da Vinci's program outcomes or impacts. Although I know there is something to this in terms of our training and background, I would not like to think that my team would shy away from qualitative work here if the clients (stakeholders) decided this was the best option. I must add, though, that the case description seems to suggest to me that test scores and quantitative measures of outcomes are important to at least some of the key stakeholders.

Finally, there has been much discussion in recent years about improving the cost-benefit and cost-effectiveness analysis of programs under evaluation. Now that I have spent some time on the other side of the fence (developing and managing programs), I am more acutely aware that we need to apply cost-benefit and cost-effectiveness concepts more routinely to our evaluation plans as well. So in the end, not only did I assume there might be human resource and financial constraints in this case, I strived to

develop a plan that would allow the stakeholders to prevent risks involved with commissioning external evaluations and would help them manage the costs of investing in this type of professional service.

JEAN A. KING: For me the Bunche–Da Vinci scenario raises the definitional specter that continues to haunt our field: exactly what activities the portmanteau concept of program evaluation envelopes. Baizerman and others (2002) state that evaluation capacity building (ECB) is its own form of practice, distinct from evaluation. Reading this chapter's analysis, I tend to agree. ECB will not work in many settings, and if it is done poorly, it may reinforce participants' negative attitudes toward evaluation—hence, the importance of situational analysis and detailed preparation prior to launching a project. To my mind, it is better not to take a contract than to proceed and fail.

This leads me to a more general point: I want to make clear that my evaluation practice includes more than ECB projects. Because I work primarily with organizations that have little money for evaluation—school districts and nonprofit organizations—my experience over time has led me to participatory methods and capacity building, first because staff members can make time for a process they believe is valuable and, second, because, by building internal capacity, the evaluation process may survive longer than a single funded project.

Students in the courses I teach quickly learn that the best answer to many procedural questions in evaluation is, "It depends," because it usually does. Alkin and Christie rightly note that a key challenge in responding to this scenario (in contrast to an actual situation) is that we were forced to make assumptions in the absence of real information regarding the "it depends" decisions. The assumptions I made stem from my capacity-building experiences in three organizations: an urban high school with an array of social and instructional problems, a large school district with constraining resources and diversity issues, and a long-time social service agency facing major changes in staff and clientele. My experience validates the opposite of what Alkin and Christie suggest: that active involvement by personnel in the Bunche–Da Vinci case or in similar situations is unlikely.

On the contrary, in seemingly desperate situations—but with the right leadership and commitment—program evaluation can become a mechanism for addressing what may seem to be the overwhelming nature of the context. Rather than adding one more item to a long to-do list, it can generate information that allows staff to tackle critical areas of practice, quickly harvest so-called low-hanging fruit, and develop strategies for long-term improvement. They can generate visible evidence that hope lives in this setting—that people can actively do something to change the circumstances both in which they work and in their clients' lives. Is this easy to do? Of course not. The phrase "with the right leadership and commitment" points to at least two

variables that, in my experience, are a necessary condition for success, one or both of which unfortunately are often lacking in such settings. My commitment is to identify places where the process may work long term, hoping in so doing to sustain long-term improvements in the organization and in the lives of children and participants.

References

- Alkin, M., and Christie, C. "An Evaluation Theory Tree." In M. Alkin, *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks, Calif.: Sage, 2004.
- Baizerman, M., Compton, D. W., and Stockdill, S. H. "New Directions for ECB." In D. W. Compton, M. Baizerman, and S. H. Stockdill (eds.), *The Art, Craft, and Science of Evaluation Capacity Building*. New Directions for Evaluation, no. 93. San Francisco: Jossey-Bass, 2002.
- Christie, C. A. "What Guides Evaluation? A Study of How Evaluation Practice Maps on to Evaluation Theory." In C. A. Christie (ed.), *The Practice-Theory Relationship in Evaluation*. New Directions for Evaluation, no. 97. San Francisco: Jossey-Bass, 2003.
- Cousins, J. B., and Whitmore, E. "Framing Participatory Evaluation." In E. Whitmore (ed.), *Understanding and Practicing Participatory Evaluation*. New Directions in Evaluation, no. 80. San Francisco: Jossey-Bass, 1998.
- Donaldson, S. I., and Christie, C. A. "The 2004 Claremont Debate: Lipsey vs. Scriven. Determining Causality in Program Evaluation and Applied Research: Should Experimental Evidence Be the Gold Standard?" *Journal of Multidisciplinary Evaluation*, forthcoming.
- Fetterman, D. M. "Empowerment Evaluation: An Introduction to Theory and Practice." In D. M. Fetterman, S. J. Kartarian, and A. Wandersman (eds.), *Empowerment Evaluation*. Thousand Oaks, Calif.: Sage, 1996.
- House, E. R. *Professional Evaluation: Social Impact and Political Consequences*. Thousand Oaks, Calif.: Sage, 1993.
- House, E. R. "Stakeholder Bias." In C. A. Christie (ed.), *The Practice-Theory Relationship in Evaluation*. New Directions for Evaluation, no. 97. San Francisco: Jossey-Bass, 2003.
- House, E. R., and Howe, K. R. *Values in Evaluation and Social Research*. Thousand Oaks, Calif.: Sage, 1999.
- Patton, M. Q. *Utilization-Focused Evaluation*. (3rd ed.) Thousand Oaks, Calif.: Sage, 1997.
- Preskill, H., and Torres, R. T. *Evaluative Inquiry for Learning in Organizations*. Thousand Oaks, Calif.: Sage, 1999.
- Scriven, M. "Reflections." In M. Alkin (ed.), *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks, Calif.: Sage, 2004.
- Shadish, W., Cook, T. and Leviton, L. *Foundations of Program Evaluation: Theories of Practice*. Thousand Oaks, Calif.: Sage, 1991.
- Stake, R. E. *Standards-Based and Responsive Evaluation*. Thousand Oaks, Calif.: Sage, 2003.
- Stufflebeam, D. L. "The CIPP Model for Evaluation." In T. Kellaghan and D. L. Stufflebeam (eds.), *The International Handbook of Educational Evaluation*. Norwood, Mass.: Kluwer, 2003.
- Yin, R. *Case Study Research: Design and Methods*. Thousand Oaks, Calif.: Sage, 1989.

MARVIN C. ALKIN is an emeritus professor in the Social Research Methodology Division in the Graduate School of Education and Information Studies at University of California, Los Angeles.

CHRISTINA A. CHRISTIE is an assistant professor and associate director of the Institute of Organizational and Program Evaluation Research at Claremont Graduate University.